Content modelling for human action detection via multidimensional approach

ABSTRACT

Video content analysis is an active research domain due to the availability and the increment of audiovisual data in the digital format. There is a need to automatically extracting video content for efficient access, understanding,browsing and retrieval of videos. To obtain the information that is of interest and to provide better entertainment, tools are needed to help users extract relevant content and to effectively navigate through the large amount of available video information. Existing methods do not seem to attempt to model and estimate the semantic content of the video. Detecting and interpreting human presence,actions and activities is one of the most valuable functions in this proposed framework. The general objectives of this research are to analyze and process the audio-video streams to a robust audiovisual action recognition system by integrating, structuring and accessing multimodal information via multidimensional retrieval and extraction model. The proposed technique characterizes the action scenes by integrating cues obtained from both the audio and video tracks. Information is combined based on visual features (motion,edge, and visual characteristics of objects), audio features and video for recognizing action. This model uses HMM and GMM to provide a framework for fusing these features and to represent the multidimensional structure of the framework. The action-related visual cues are obtained by computing the spatio temporal dynamic activity from the video shots and by abstracting specific visual events. Simultaneously, the audio features are analyzed by locating and compute several sound effects of action events that embedded in the video. Finally, these audio and visual cues are combined to identify the action scenes. Compared with using single source of either visual or audio track alone, such combined audio visual information provides more reliable performance and allows us to understand the story content of movies in more detail. To compare the usefulness of the proposed framework, several experiments were conducted and the results were obtained by using visual features only (77.89% for precision;72.10% for recall), audio features only (62.52% for precision; 48.93% for recall)and combined audiovisual (90.35% for precision; 90.65% for recall).

Keyword: Audiovisual; Semantic; Multidimensional; Multimodal; Hidden markov model